Ann. Inst. Statist. Math.
Vol. 51, No. 4, 749–753 (1999)

ON A CHARACTERIZATION OF THE GAMMA DISTRIBUTION: THE INDEPENDENCE OF THE SAMPLE MEAN AND THE SAMPLE COEFFICIENT OF VARIATION

TEA-YUAN HWANG[1] AND CHIN-YUAN HU[2]

[1] *Institute of Statistics, National Tsing Hua University, Hsinchu 30043, Taiwan, R.O.C.*
[2] *Department of Business Education, National Changhua University of Education, Changhua 50058, Taiwan, R.O.C.*

(Received January 26, 1996; revised June 24, 1998)

Abstract. Let $n \geq 3$ and let $X_1, \ldots, X_n$ be positive i.i.d. random variables whose common distribution function f has a continuous p.d.f. Using earlier work of the present authors and a method due to Anosov for solving certain integro-functional equations, it is shown that the independence of the sample mean and the sample coefficient of variation is equivalent to that f is a gamma function. While the proof is of methodological interest, this conclusion can also be arrived at without any assumptions by appealing to the Laplace-Stieltjes transform, as in the Concluding Remark (Section 3).

Key words and phrases: Characterization, gamma distribution, coefficient of variation.

1. Introduction and main result

Two of the statistics most often used in both theory and applied work are the sample mean $\bar{X}_n$ and the sample standard deviation S_n. It is also well-known that the independence of $\bar{X}_n$ and S_n (based on a random sample) characterizes the normal population. Various results characterizing the parent distribution through various properties of statistics can be found in Kagan *et al.* (1973)—abbreviated in what follows as Kagan *et al.* (1973), Johnson and Kotz (1970), Lukacs and Laha (1964) and the references therein. However, it seems that characterization problems based on the properties of the coefficient of variation $V_n = S_n/\bar{X}_n$ have seldom been studied. This may be the reason why V_n is not used often.

In this context, we establish

THEOREM. *Let $n \geq 3$ and let $X_1, \ldots, X_n$ be n positive i.i.d. random variables with their common distribution function having a probability density function $f(x)$. Then the independence of the sample mean $\bar{X}_n$ and the sample coefficient of variation $V_n = S_n/\bar{X}_n$ is equivalent to that f is a gamma density.*

The above could possibly be misunderstood as an immediate consequence of a known theorem such as Theorem 6.2.9, p. 202, Section 6.2 of Kagan *et al.* (1973). As a matter of fact, their theorem needs the condition of the constancy of regression of V_n on $\bar{X}_n$ which is weaker than that of the independence of V_n and $\bar{X}_n$, but, on the other hand, presupposes the existence of some moments of the X_i's. Our theorem holds without any conditions on the moments of the X_i's.

Three lemmas used to prove the main theorem will be presented in Section 2; the proofs for those lemmas are partially based on that of Anosov's theorem (1964) and the recent result of Hwang and Hu (1994). Section 3 gives the proof of the main theorem.

The conclusion of the theorem can be arrived at without any assumptions, by appealing to the Laplace-Stieltjes transform, as pointed out by a referee—see the Concluding Remark at the end of Seciton 3. Our proof is given here, for the less general situation, as of methodological interest.

2. Three lemmas

For convenience of citation in the proof of the three lemmas used for proving the main theorem, we cite certain recent results obtained by Hwang and Hu (1994) as follows:

Define a non-linear transformation $(x_1, \ldots, x_n) \to (t_1, \ldots, t_{n-2}, \bar{x}_n, v_n)$, where

$$t_i = \left[\frac{n-i+1}{(n-1)(n-i)}\right]^{1/2} \cdot \left[\frac{x_i - \bar{x}_n}{s_n} + \frac{1}{n-i+1}\sum_{k=1}^{i-1}\frac{x_k - \bar{x}_n}{s_n}\right], \quad 1 \le i \le n-2 \tag{2.1}$$

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad v_n = s_n/\bar{x}_n,$$

the summation in (2.1) is taken as zero for $i = 1$, and s_n is the standard deviation of $x_1, \ldots, x_n$. Then, Theorem 2.2 of Hwang and Hu (1994) gives

$$\begin{aligned}
\lambda_i(t) &= \left[\frac{(n-i)(n-1)}{n-i+1}\right]^{1/2} \cdot t_i - \sum_{k=1}^{i-1}\left[\frac{n-1}{(n-k)(n-k+1)}\right]^{1/2} \cdot t_k, \quad 1 \le i \le n-2 \\
\lambda_{n-1}(t) &= -\left[\frac{(n-1)\cdot f_{n-2}}{2}\right]^{1/2} - \sum_{k=1}^{n-2}\left[\frac{n-1}{(n-k)(n-k+1)}\right]^{1/2} \cdot t_k, \\
\lambda_n(t) &= \left[\frac{(n-1)\cdot f_{n-2}}{2}\right]^{1/2} - \sum_{k=1}^{n-2}\left[\frac{n-1}{(n-k)(n-k+1)}\right]^{1/2} \cdot t_k,
\end{aligned} \tag{2.2}$$

where $\lambda_i(t) = (x_i - \bar{x}_n)/s_n$ and $f_{n-2} = 1 - t_1^2 - \cdots - t_{n-2}^2$. Thus, we have

$$\sum_{i=1}^{n}\lambda_i(t) = 0, \qquad \sum_{i=1}^{n}\lambda_i^2(t) = n-1 \tag{2.3}$$

$$x_i = \bar{x}_n[v_n \cdot \lambda_i(t) + 1], \qquad 1 \le i \le n. \tag{2.4}$$

We have then:

LEMMA 2.1. *Let $n \geq 3$ and let $X_1, \ldots, X_n$ be n positive i.i.d. random variables having a p.d.f. $f(x)$, and let $\bar{X}_n$ and V_n be the sample mean and the sample coefficient of variation. Then, the joint p.d.f. $f(x,v)$ of $(\bar{X}_n, V_n)$ is*

$$c_n \cdot x^{n-1} \cdot v^{n-2} \cdot \int \cdots \int_{B_{v,n}} \pi_{i=1}^n f(x(v\lambda_i(t)+1)) d\mu(t) \tag{2.5}$$

for $x > 0$ and $0 < v < \sqrt{n}$, and zero otherwise, where $c_n = n!\sqrt{n}(n-1)^{(n-1)/2}$, the functions $\lambda_i(t)$ are defined as in (2.2), $d\mu(t) = f_{n-2}^{-1/2} dt_1, \ldots, dt_{n-2}$ and the set $B_{v,n}$ depends on v, for $0 < v < \sqrt{n}$, as follows:

$$B_{v,n} = \left\{ t : \begin{array}{l} \max\left\{ \dfrac{-\sqrt{n}}{(n-1)v}, -1 \right\} \leq t_1 \leq \dfrac{-1}{n-1} \\ \max\left\{ \left(\dfrac{n-k+2}{n-k} \right)^{1/2} \cdot t_{k-1}, -f_{k-1}^{1/2} \right\} \leq t_k \leq -\dfrac{f_{k-1}^{1/2}}{n-k} \\ 2 \leq k \leq n-2, \end{array} \right\}, \tag{2.6}$$

and $f_i = 1 - t_1^2 - \cdots - t_i^2$, $1 \leq i \leq n-2$.

LEMMA 2.2. *Under the condition of Lemma 2.1, assume that $\bar{X}_n$ and V_n are independent. Then, the p.d.f.'s of $\bar{X}_n$ and V_n are respectively given by*

$$f_{\bar{X}_n}(x) = a_n \cdot x^{n-1} \cdot [f(x)]^n, \qquad x > 0 \tag{2.7}$$

and zero otherwise, where a_n is a normalizing constant;

$$f_{V_n}(v) = b_n \cdot v^{n-2} \cdot \int \cdots \int_{B_{v,n}} \pi_{i=1}^n f(v\lambda_i(t)+1) d\mu(t), \qquad 0 < v < \sqrt{n} \tag{2.8}$$

and zero otherwise, where b_n is a normalizing constant, and the $\lambda_i(t)$ and $B_{v,n}$ are defined as in (2.2) and (2.6) respectively.

PROOF. Let $f_{\bar{X}_n}(x)$ and $f_{V_n}(v)$ be the p.d.f.'s of $\bar{X}_n$ and V_n respectively; it follows from Lemma 2.1 and the independence of $\bar{X}_n$ and V_n that the joint p.d.f. of $(\bar{X}_n, V_n)$ must be equal to the product of their densities,

$$f_{\bar{X}_n}(x) \cdot f_{V_n}(v) = c_n \cdot x^{n-1} \cdot v^{n-2} \cdot \int \cdots \int_{B_{v,n}} \pi_{i=1}^n f(x(v\lambda_i(t)+1)) d\mu(t)$$

for all $x > 0$ and $0 < v < \sqrt{n}$. We see that $f_{\bar{X}_n}(1) \neq 0$, since otherwise the right side of the equation would vanish for all $0 < v < \sqrt{n}$, which is impossible. Set $x = 1$ in the equation. Then, we obtain the expression for $f_{V_n}(v)$ as given in (2.8).

Now substituting (2.8) in the equation, and dividing both sides of the equation by v^{n-2}, we get

$$f_{\bar{X}_n}(x) \cdot b_n \int \cdots \int_{B_{v,n}} \pi_{i=1}^n f(v\lambda_i(t) + 1) d\mu(t)$$
$$= c_n \cdot x^{n-1} \cdot \int \cdots \int_{B_{v,n}} \pi_{i=1}^n f(x(v\lambda_i(t) + 1)) d\mu(t).$$

Note that if the variable v is in the neighborhood of origin, say $0 < v < \sqrt{n}/(n-1)$, then the domain of integration $B_{v,n}$ is independent of v; first by using this fact, and then letting $v \to 0^+$ in the new equation, we obtain the expression for $f_{\bar{X}_n}(x)$ as given in (2.7). Thus, we have established Lemma 2.2. □

LEMMA 2.3. *(An integro-functional equation) Under the conditions of Lemma 2.2, the following integro-functional equation holds:*

$$\tag{2.9} \int_{B_{v,n}} \pi_{i=1}^n f(x \cdot (v\lambda_i(t) + 1)) d\mu(t) = c_n \cdot [f(x)]^n \cdot \int_{B_{v,n}} \pi_{i=1}^n f(v\lambda_i(t) + 1) d\mu(t)$$

for all $x > 0$ and $0 < v < \sqrt{n}$; where $c_n > 0$, depending on n, is a constant, the $\lambda_i(t)$ and $B_{v,n}$ are defined as in (2.2) and (2.6) respectively, and $t = (t_1, \ldots, t_{n-2})$ is a point on the $(n-2)$-dimensional set $B_{v,n}$. In particular, if v is in the neighborhood of the origin, then the domain of integration $B_{v,n}$ is independent of the variable v, that is, $B_{v,n}$ is replaceable by B_n in (2.9) for $0 < x$ and $0 < v < \sqrt{n}/(n-1)$, where

$$\tag{2.10} B_n = \left\{ t : \begin{array}{l} -1 \le t_1 \le \dfrac{-1}{n-1} \\ \max\left\{ \left(\dfrac{n-k+2}{n-k}\right)^{1/2} \cdot t_{k-1}, -f_{k-1}^{1/2} \right\} \le t_k \le -\dfrac{f_{k-1}^{1/2}}{n-k} \\ 2 \le k \le n-2, \end{array} \right\},$$

and $f_i = 1 - t_1^2 - \cdots - t_i^2$, $1 \le i \le n-2$.

PROOF. This lemma follows immediately from Lemmas 2.1 and 2.2 and the independence of $\bar{X}_n$ and V_n. □

3. Proof of the Theorem

It is easy to show that the sample mean $\bar{X}_n$ and the sample coefficient of variation $V_n = S_n/\bar{X}_n$ are independent if the parent population is gamma. This fact follows immediately from Lemma 2.1 by taking the first relation in (2.3) into account.

Conversely, it follows from Lemma 2.3 that $f(x)$ satisfies the integro-functional equation (2.9) for all $x > 0$ and $0 < v < \sqrt{n}$ if $\bar{X}_n$ and V_n are independent, thus it is of the same form as considered by Anosov (1964), as reproduced in Kagan *et al.* (1973), pp. 143–148, Section 4.9. The roles of t, s, ϕ used there are played here by $\bar{x}$, $\bar{x}\cdot v$ and $\boldsymbol{t}$; and, instead of the range of integration $[0, 2\pi]$ for ϕ, we have B_v for $\boldsymbol{t}$, this set not depending on v if $0 < v < \sqrt{n}/(n-1)$. If $u := \log f$ on a fixed maximal open interval $I \subset (0, \infty)$ where $f > 0$, then defining $L_{v,t}u(x)$ and proceeding essentially as in proving Anosov's theorem, we conclude that $u(x) = A + B\log x + Cx$ for $x \in I$, then that $I = (0, \infty)$, and finally that f is a gamma density.

Concluding Remark. A referee has pointed out that, while our proof is of methodological interest, the conclusion of the main result can be arrived at, under no assumptions whatever, by appealing to Laplace-Stieltjes transform. For simplicity, write (respectively) S, V for S_n, V_n and let

$$T = \sum X_j = n\bar{X} \quad \text{and} \quad S^2 = \sum (X_j - \bar{X})^2; \qquad V = S/\bar{X}.$$

Let F be the common d.f. of the X_j and ϕ their LST: $\phi(t) = \int_0^\infty e^{-tX} dF(x)$, $t > 0$. If $\bar{X}$ and V are independent v.r.'s, then, for every $t > 0$, the r.v.'s $T^2 e^{-tT}$ and V^2 are bounded, positive, independent r.v.'s and so

$$E(T^2 e^{-tT} \cdot V^2) = E(T^2 e^{-tT}) \cdot EV^2 (< \infty)$$

easily leading to

$$E\left(\sum X_j^2 e^{-tT}\right) = \text{const.} E\left(\sum_{j \neq k} X_j X_k e^{-tT}\right),$$

whence

$$\phi''(t) \cdot \phi(t) = c(\phi'(t))^2$$

on $(0, \infty)$. The only (non-trivial probabilistic) solutions are given by $\phi(t) = (1 - at)^{-b}$ for some $a, b > 0$. Hence F has gamma p.d.f. by inversion formula for LST.

Acknowledgements

The authors are grateful to a referee for the Remark in the end of the previous Section and for suggestions which have greatly improved the paper.

References

Anosov, D. V. (1964). On an integral equation arising in statistics, *Vestnik Leningrad University*, **7**, 151–154.

Hwang, T. Y. and Hu, C. Y. (1994). On the joint distribution of studentized order statistics, *Ann. Inst. Statist. Math.*, **46**, 165–177.

Johnson, N. L. and Kotz, S. (1970). *Continuous Univariate Distributions-1*, Houghton Miffin Co., Boston.

Kagan, A. M., Linnik, Y. V. and Rao, C. R. (1973). *Characterization Problems in Mathematical Statistics* (translated by B. Ramachandran), Wiley, New York.

Lukacs, E. and Laha, R. G. (1964). *Applications of Characteristic Functions*, Hafner, New York.